

Manjon-Antolin, Miguel C.; Arauzo-Carod, Josep Maria

**Conference Paper**

## Locations and Relocations: Modelling, Determinants, and Interrelations

46th Congress of the European Regional Science Association: "Enlargement, Southern Europe and the Mediterranean", August 30th - September 3rd, 2006, Volos, Greece

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Manjon-Antolin, Miguel C.; Arauzo-Carod, Josep Maria (2006) : Locations and Relocations: Modelling, Determinants, and Interrelations, 46th Congress of the European Regional Science Association: "Enlargement, Southern Europe and the Mediterranean", August 30th - September 3rd, 2006, Volos, Greece, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/118189>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# **Locations and Relocations: Modelling, Determinants, and Interrelations<sup>\*</sup>**

Miguel C. Manjón-Antolín<sup>\*\*</sup>

(Department of Economics and GRIT, Rovira i Virgili University)

and

Josep-Maria Arauzo-Carod

(Department of Economics and GRIT, Rovira i Virgili University)

*Very preliminary version*

---

<sup>\*</sup> We are grateful to J. Trullén and R. Boix for kindly providing their municipalities' database. Thanks are also due to W. Boente, T. Mora, I. Peña and seminar participants at CAEPS (University of Barcelona) and the Max Planck Institute of Economics for their useful comments to earlier drafts of this paper. We also acknowledge the institutional support of the CIDEM. This paper is part of a research project financially supported by grants SEJ2004-05860/ECON and SEJ2004-07824/ECON of the CICYT. Any errors are of course our own.

<sup>\*\*</sup> Corresponding author ([miguel.manjon@urv.net](mailto:miguel.manjon@urv.net)). Address for correspondence: Avda. Universitat 1, Reus-43204, Spain.

# **Locations and Relocations: Modelling, Determinants, and Interrelations**

## Abstract

Most studies in industrial location do not distinguish between new and relocated establishments. This paper addresses this shortcoming using data on the frequency of these events in municipalities of the same (Spanish) region (Catalonia). This enables us to test not only for differences in their determinants but also for interrelations between start-ups and relocations. Estimates from count regressions models for cross-section and panel data show that, although partial effects differ, common patterns arise in “institutional” and “neoclassical” explanatory factors. Also, start-ups and relocations are positively related.

JEL classification: C25, R30, R10

Keywords: cities, count data models, industrial location

## 1. Introduction

There is an extensive empirical literature on the determinants of industrial location.<sup>1</sup> However, most of this research implicitly assumes that start-ups are all similar. This assumption contrasts with evidence that shows that start-ups significantly differ in features such as e.g. size (Carlton 1979), technology (Frenkel 2001) and country of origin (Figueiredo et al. 2002). More importantly, these studies conclude that these differences critically affect the way location decisions are taken.

In this paper we examine yet another feature that has comparatively received less attention. Namely, the fact that some start-ups are actually relocations, i.e. businesses established in the past that at some point decide to abandon its current location and move to another one. The distinction is important because location decisions are taken on the grounds of incomplete information about the sites and so previous experiences can make a difference. Therefore, the opening of new concerns and the relocation of existing concerns are different location processes which should consequently be studied separately (Pellenbarg et al. 2002a, 2002b).

Accordingly, we analyse the determinants of industrial location distinguishing between strictly-new and relocated concerns. Also, we explore the possibility that both processes, locations and relocations, are interrelated (“attraction” and “rejection” effects). From an economic policy point of view, these are issues worth inspecting given the increasingly large amount of public funds invested in public incentive programs aiming to attract new businesses (Lee 2004). We use the establishment and the municipality as units of analysis.<sup>2</sup>

---

<sup>1</sup> See e.g. Arauzo et al. (2006) and Basile and McCann (2006) for a survey of this literature.

<sup>2</sup> More specifically, we use data from the “Registry of Industrial Establishments of Catalonia” on the establishments located in the 946 Catalan municipalities in the period 2001-2004. Catalonia is a Spanish region (NUTS-II) in the northeast of Spain whose capital is Barcelona. It has an area of 31,895 km<sup>2</sup>, a population of about 7 million people (15% of the population of Spain) and its GDP is approximately 20% of Spain’s GDP. The reason for choosing Catalonia (instead of e.g. any other Spanish region) to carry out this study was the richness of the municipality data-base kindly provided by Trullén and Boix (2004), which we complemented with data from the Catalan Institute of Statistics to construct our vector of explanatory variables.

This paper fits into the Foreign Direct Investments (FDI) literature to the extent that we share an interest in the relocation of production (see e.g. Blonigen 2006). In contrast with the typical FDI paper, however, relocations in our dataset occur within the same economic administrative region. This means that we do not have data on either cross-border activities or, for what matters, the municipality from which relocations come from. This paper also differs from previous studies on the firm/plant relocation decision within the same country in that we are mainly concerned with how the characteristics of a municipality affect the rate of occurrence of start-ups and relocations rather than with how firm/plant (and, in some studies, location) characteristics affect the choice of a particular municipality from a set of potential (re)locations.<sup>3</sup>

The closest study in this respect is that of Holl (2004a), who analyses the spatial patterns of start-ups and relocations in Portugal between 1986 and 1997. Because of the integer nature of our dependent variable, the number of new establishments in a (Catalan, Portuguese) municipality, we both resort to count data models. However, there are important differences with our study. First, she studies a larger geographical area that covers many heterogeneous regions (although she somehow controls for this). Second, she uses a set of explanatory variables that, differences in sources and definitions notwithstanding, is largely nested in our specification. Third, she does not explore alternative assumptions on the data generation process as we do. Fourth, she does not address the interrelation between start-ups and relocations.

We find that the determinants of start-ups and relocations are practically the same. These include “neoclassical” and “institutional” factors such as e.g. (dis)urbanisation and location economies (neoclassical) as well as dummies for the administrative and spatial organisation of the territory (institutional, on which Holl incidentally has no data). However, the partial derivatives of the conditional expectation of start-ups and relocations with respect to these determinants, i.e. the partial or marginal effects, differ. Hence, our results do not fully concur with Holl’s main “finding (...) that plant start-ups and relocations are not attracted by the same set of location characteristics” (p. 665). Also, we provide

---

<sup>3</sup> Studies focusing on this choice include Erikson and Wasylenko (1980), Baudewyns et al. (2000), van Dijk and Pellenbarg (2000), Brower et al. (2004), Lee (2004) and Strauss-Kahn and Vives (2005).

evidence of a positive statistical relation between the rates of occurrence of these events. That is, the likelihood that new and previously established concerns (re)locate in a particular municipality increases with the presence of relocations and star-ups, respectively, in that municipality.

The rest of the paper is organised as follows. Section 2 discusses the econometric methodology. Section 3 briefly reviews the literature on the determinants of industrial (re)location. Since we estimate reduced-form models this review will provide the theoretical foundations for the vector of explanatory variables. Section 4 contains the empirical results. We describe the data and present inferences from count regression models. Section 5 is the conclusion.

## **2. Econometric modelling**

Discrete choice models (DCM) and count data models (CDM) are the basic econometric tools in empirical studies of industrial location (Arauzo et al. 2006). One reason for this is that these models are consistent with a profit maximization framework in which firms choose the optimal location subject to standard constraints (see e.g. Becker and Henderson 2000 and Guimarães et al. 2004). However, DCM critically differ from CDM in the type of data they require and the type of inferences they provide. On the one hand, the unit of analysis in DCM is the firm or the establishment and the main concern are how certain characteristics of this unit (size, sector, etc.) and/or the chosen territory (population, infrastructures, etc.) affect location decisions. On the other hand, the unit of analysis in CDM is geographical (municipality, county, province, region, etc.) and the factors that may affect location decisions refer accordingly to the territory.

In light of these features one could argue that DCM have an advantage over CDM because they may account for both firm and spatial factors. However, there are other issues worth considering when it comes to selecting a model for our location study. One should be aware, for example, that computation of the likelihood function in DCM is cumbersome when the number of alternatives (i.e. locations) is large. Moreover, the set of alternatives in DCM only includes those locations effectively chosen, since the rest do not contribute to the likelihood function. These drawbacks can turn CDM into our preferred

specification, especially whenever is possible to recover the parameter estimates of DCM from the estimates of CDM and/or the sample contains a substantial number of locations where not a single business started operations in the period of analysis (Guimarães et al. 2003). Computational burden is not an issue in CDM and zero observations not only contribute to the likelihood function but provide interesting insights about the data generation process (Cameron and Trivedi 1998).

CDM seem therefore particularly useful for investigations using highly geographically disaggregated data. We can mention, among others, the urban studies of Holl (2004b) on Portuguese municipalities, Holl (2004c) on Spanish municipalities and Arauzo and Manjon (2004) and Arauzo (2005) on Catalan municipalities. Yet there are studies that apply CDM to larger geographical units, such as e.g. Becker and Henderson (2000), List and McHome (2000) and Guimarães et al. (2004) on US counties and Papke (1991) on US states.<sup>4</sup>

All these studies, however, do not distinguish between strictly-new concerns and relocations. To the best of our knowledge, the only previous study that makes such a distinction and uses CDM is Holl (2004a). She analyses the determinants of Portuguese plant start-ups and plant relocations using Fixed-Effects estimators for Poisson and Negative Binomial models to control for unobserved municipality-specific heterogeneity. In this paper we essentially follow the same approach. However, we explore alternative specifications to cope with the distinct characteristics of our data. Namely, “excess of zeros” and overdispersion, possible discrepancy between the period of occurrence (exposure) and the period of observation, and dependence between the events of interest. Next we discuss these in detail.<sup>5</sup>

## 2.1 Excess of zeros and overdispersion

As Figure 1 shows, there are more zeros in the histograms of start-ups and relocations than the Poisson density (with unitary mean) predicts. This largely disqualifies the standard Poisson regression model as a suitable specification for our data. Also, descriptive statistics in Panel A of Table 1 indicate that the

---

<sup>4</sup> CDM are also very popular in the FDI literature —see e.g. Basile and McCann (2006) and Blonigen (2006).

<sup>5</sup> The statistical foundations of most of what follows can be found in Cameron and Trivedi (1998, 2001).

conditional variance of our processes exceeds their conditional mean. To see this notice that although the conditional variance in Poisson regression models is generally smaller than the non-conditional variance, the conditional expectation should not differ considerably from the sample mean (as long as the model has a constant term). Since “equidispersion”, i.e. equality of conditional variance and mean, is one the main assumptions of the Poisson regression model, its rejection further supports the need for less restrictive, more efficient models. One of such models are the so-called “mixture models”.

[Insert Figure 1 about here]

Mixture models essentially differ from the Poisson regression model in that we introduce a stochastic term  $\xi$  in the conditional mean function of the dependent variable, usually in multiplicative form. In standard notation,  $E(y|x, \xi) = \exp(x\beta)\xi = \mu\xi$ . Within this basic framework, two large classes of CDM arise depending on whether  $\xi$  is considered a continuous or discrete variable. (i) In continuous mixtures  $\xi$  has a natural interpretation as an individual random effect that allows for unobserved heterogeneity in the model. That is,  $\xi$  aims to account for characteristics of the municipality that are not observed by the researcher (e.g. the “business climate”) and/or differences in municipalities beyond those captured by the explanatory variables. Assuming that  $\xi$  has a Gamma distribution with unitary mean and constant variance  $\alpha$ , for example, leads to one of the most popular continuous mixtures: the Negative Binomial Model (NBM). (ii) In finite mixtures  $\xi$  allows for the existence of a discrete number of heterogeneous groups in the population of interest. In the simplest case this amounts to assume that the population consists of two groups: municipalities in which there are not nor will be new/relocated concerns (e.g. because they are banned by environmental regulations) and municipalities in which there are not but might be new/relocated concerns (i.e. in principle there is nothing that prevents this to happen). To construct the finite mixture model this binary-form of heterogeneity is parameterised using e.g. the logistic transformation and the resulting logit model for the probability of zero entrants in the municipality is mapped into a count model that now only accounts for the positive values of the dependent variable. The Zero Inflated Poisson Model (ZIPM) uses the Poisson model to this end, whereas using NBM generates the Negative Binomial Inflated Model (NBIM).



This is a convenient way to extend the basic Poisson regression model because mixing strategies based on a multiplicative term induce both excess of zeros and overdispersion in the distribution (Mullahy 1997). It is important to keep in mind, however, the way in which each model brings this result about. NBIM controls for heterogeneous municipalities but assumes the same data generation process for zeros and positive outcomes. In contrast, ZIPM distinguishes two regimes in the data generation process. The downside is that ZIPM does not account for overdispersion in the positive set. NBIM addresses this downside, though it is much less parsimonious than NBM and ZIPM. The practical problem is that we cannot easily discern which of these mechanisms is ultimately responsible for the observed excess of zeros and overdispersion.<sup>6</sup>

Several tests may help us to select the appropriate model for our data. First, we can test for overdispersion (i.e. rejection of the Poisson model) using a LR test on the  $\alpha$ -parameter of the variance of  $y$  in NBM,  $\mu + \alpha\mu$ . This test is denoted by “LR  $\alpha$  Test” in the tables of results presented in Section 4. A LR-type test can also be implemented to compare the nested structures of the ZIPM and the NBIM. This test is denoted by “LR ZIPM-NBIM Test” in the tables of results. Lastly, the “Vuong Test” (also reported) provides a non-nested testing procedure used to discriminate between Poisson and Negative Binomial models and their respective analogous inflated, i.e. ZIPM and NBIM. In addition to these vis-à-vis tests between models, it is useful to analyse some statistics regarding each model. These include the value of the log-likelihood function (denoted by “Log L” in the tables of results), the Likelihood-Ratio test for the joint significance of the model (“LR Joint Test”) and the  $\chi^2$  goodness-of-fit test (“GoF Test”).

## 2.2 Exposure

---

<sup>6</sup> This identification problem becomes even more involved if we consider that failure of the “independence of events” assumption in the Poisson process also causes overdispersion (Cameron and Trivedi 2001: 337). In the context of industrial location studies this assumption means that the likelihood that an establishment locates in a particular municipality is independent of other establishments being located there. This may sound as a restrictive assumption, but for simplicity we should not pursue this issue here. Notice, however, that the discussion on the dependence between start-ups and relocations below partially addresses this issue.

Another characteristic of the data that deserves attention concerns the definition of the exposure. Ideally, the period of observation should be the same as the period of occurrence of the events. However, this is often not the case. Our data, for example, are annually recorded. But this does not necessarily mean that the exposure period is annual. In fact, there is no economic or legal reason why the rate of occurrence of start-ups and relocations should be calculated on a yearly basis rather than over the available four-year period or, in general, over any other period. We attempt to deal with this indeterminacy by considering alternative definitions of the exposure period.

We initially assume that the periods of occurrence and observation are the same. Accordingly, the dependent variables are the number of start-ups and relocations reported over the period 2001 to 2004 and the explanatory variables are calculated as period-means. We report these estimates in Table 2. Later we assume that the period of occurrence corresponds to the period used by the statistical sources, in this case the “Registry of Industrial Establishments of Catalonia”. This lead us to examine two additional questions.

The first is whether the data generation process governing the yearly rates of occurrence of start-ups and relocations is the same every year. To answer this question we could simply compare the estimates from different sample years. However, our results would be distorted by the fact that many of our explanatory variables are only available for the year 2001 (see Table 1). Consequently, in Table 3.A we present estimates for the year 2001 and use as a benchmark those obtained from pooling the data (reported in Table 3.B). Since all these proposals imply cross-section data structures NBM, ZIPM and NBIM are suitable estimation settings.

The second is whether the yearly rates of occurrence of start-ups and relocations are independent. To answer this question we calculated the covariance matrix for the year vector of Pearson-residuals from the pooled Poisson regression model. As Hausman et al. (1984) argue, if the assumption of time independence holds the resulting  $4 \times 4$  matrix should have small values in the off diagonal elements and cross-section estimates would provide valid inferences. The estimated correlations are indeed practically

negligible, with highest values of order 0.3 but typically around 0.02. We thus find evidence supporting the time independence property in our data.<sup>7</sup>

Still, it may be interesting to take full advantage of the panel structure of our data set (e.g. because of efficiency gains). As discussed in the previous subsection, NBM and NBIM enable us to control for unobserved heterogeneity in cross section data. However, they are not Poisson models. Panel data estimators also control for unobserved heterogeneity, but they can do it maintaining the assumption that the data is Poisson distributed. These estimates may therefore be useful to assess the robustness of the conclusions extracted from negative binomial models using cross-section data. On the other hand, panel data estimators based on the Poisson distribution impose equidispersion. As in the cross-section case, Negative Binomial specifications may cope with this.

We accordingly report results from Poisson and Negative Binomial fixed- and random-effects estimators (hereafter FE and RE, respectively) in Table 4. The issue, though, is not whether the latent effects are considered fixed or random but what is their stochastic relation with the covariates, for it is this relation what determines the statistical properties of the estimators. The RE estimator is not consistent if the (strictly exogenous) covariates are correlated with the effects, whereas the FE estimator is consistent anyway. Moreover, zero correlation between covariates and latent effects renders the RE estimator efficient. These properties match the general conditions of the “Hausman test” (reported in Table 4), which may thus help us to choose between FE and RE. Also worth considering in this decision is the fact that time-invariant explanatory variables are not identified in the conditional (on a sufficient statistic) maximum likelihood framework that sustains the FE estimator. Since a good deal of our explanatory variables is of this kind, we have decided to report both FE and RE estimates. Nevertheless, the latter should be interpreted with care in those cases where the null hypothesis of independence between covariates and effects is rejected.

---

<sup>7</sup> If the time independence assumption does not hold panel data estimators are needed to control for the serial correlation induced by the presence of individual effects. Cross-section estimators cannot “distinguish (...) between true time independence versus apparent dependence due to the unobserved heterogeneity of the individual units”

## 2.3 Dependence

Our last concern with the data relates to the likely dependence between the events of interest. We have argued that the opening of new and relocated establishments are different location processes. However, so far we have implicitly assumed that they are totally independent. Here we address this assumption.

In general, a positive relationship between start-ups and relocations may arise if, for example, start-ups interpret the occurrence of relocations in a municipality as a sign that this municipality is a better location than those from where relocations moved from (which would consequently be judged as “bad locations”). The relation would also be positive if, for example, relocated businesses happen to be main suppliers of start-ups that have decided to move closer to their consumers to save costs, improve contacts, etc.. However, a negative relationship may also be possible if, for example, start-ups and relocations see each other as powerful competitors that may throw them out of the local market. Under these circumstances they may sensibly opt for installing their establishments in other locations.

Admittedly, these examples do not constitute a solid theoretical argument. Still, they suggest that an “attraction effect” (positive) and a “rejection effect” (negative) between start-ups and relocations may exist. It is therefore a matter of empirical research to verify this tenet. The sample correlation between the number of (pooled) start-ups and relocations is 0.73, thus casting doubts on the independence assumption. However, we clearly need to go beyond descriptive statistics to obtain sounded evidence. The problem is that since these effects are not mutually exclusive, i.e. they may occur simultaneously in a municipality, we cannot econometrically identify them using CDM and the municipality as unit of analysis. What we can and will do is to test whether (on average) one dominates the other.

To implement this test in our models we need a specification that separates the effects that start-ups may cause on relocations from those that relocations may cause on start-ups. The first effect takes place when

---

(Hausman et al. 1984: 911). This is the assumption implicitly made by Holl (2004a). Notice, however, that she does not provide supportive statistical evidence.

the likelihood that businesses located in a particular municipality move to another one is, *ceteris paribus*, positively/negatively affected by the observation that that location has been chosen for starting up new businesses. The second effect takes place when the likelihood that new businesses start-up activities in a particular municipality is, *ceteris paribus*, positively/negatively affected by the observation that this municipality has been chosen for relocating existing concerns. Our setting looks then like a simultaneity framework in which positive (negative) signs in the coefficients indicate attraction (rejection).

Notice, however, that the previous examples and the proposed specification suggest that the interrelation between start-ups and relocations is more of a sequential than a simultaneous nature. That is, the timing of the underlying game seems to consist of two stages: observation and decision. Accordingly, we will proceed in an analogous way as when we assumed that the periods of occurrence and observation are the same. The difference is that now we will restrict the sample to the 2002 to 2004 period and will include the number of start-ups and relocations in 2001 among the determinants of relocations and start-ups, respectively. The coefficients of these variables will provide evidence of (the mean dominance of) “attraction” and/or “rejection” effects in a typical municipality. These are reported in Table 5.

### **3. The determinants of industrial (re)location**

Having discussed the econometric modelling we turn now to the selection of the explanatory variables. To this end, this section briefly reviews the theoretical and empirical literature on the determinants of industrial (re)location. This will provide guidelines for the construction of our specifications as well as reveal the differences and similarities between our investigation and previous related studies.

#### **3.1 Theories**

The location of economic activity has been analysed from a wide range of theoretical perspectives. None of them, however, have dedicated much effort to investigate the idiosyncrasies of relocations. As Brower et al. (2004: 336) point out, “[r]elocation theories are hardly applied and are often treated as a special case of location theories”. In any case, a thorough discussion of each theory is clearly beyond the scope of this paper. For our purposes it suffices to outline their principal tenets. In particular, we follow Hayter

(1997) in distinguishing three main approaches in location theory: neoclassical, behavioural and institutional.

The decision setting in the neoclassical theory involves rational agents choosing optimally a site among a set of finite alternatives. Hence, in this framework the main determinants of industrial location are those affecting the expected benefits derived from the decision to locate in a particular site. These include, for example, transportation and labour costs, external economies and market size. Most of the studies discussed in the previous section are largely based on these (constrained) profit-maximisation or cost-minimising strategies —see e.g. Frenkel (2001) and Guimarães et al. (2003, 2004).

As for the behavioural theory, it stems essentially from the same decision setting as that of the neoclassical theory. However, it calls into question the assumptions of rationality and perfect information. This means that agents have limited knowledge and take their location decisions in a world of uncertainty. Unlike the neoclassical approach, which places great emphasis on “external” factors, the behavioural approach stresses the importance of “internal” (size, age, etc.) and “entrepreneurial” (previous experience, residence, etc.) factors in the location decision. Supportive evidence shows, for example, that large firms tend to consider larger sets of alternatives than small firms (Arauzo and Manjon 2004) and that entrepreneurs are more likely to choose locations near her/his residence (Figueiredo et al. 2002).

Lastly, the institutional theory disagrees with the notion held by neoclassical and behavioural theories that firms are isolated agents. In fact, this theory notices that is quite the opposite: firms operate within a network of clients, suppliers, competitors, trade unions, regional systems, governments, etc.. The environment matters and should consequently be taken into account when modelling location decisions. Institutional theory advocates accordingly for paying more attention to issues such as e.g. wages, unionisation and regulations. Carlton (1983), Papke (1991) and List and McHome (2000), for example, provide empirical evidence on some of these issues.

All in all, these theories provide a solid analytical framework for a variety of research questions in industrial location. But are they that useful for studying relocations? Pellenbarg et al. (2002a: 11) contend that location theories do “provide the theoretical background for studies of firm relocation”. This is also what implicitly maintain van Dijk and Pellenbarg (2000), Brower et al. (2004) and Holl (2004a), for they all refer to neoclassical, behavioural and institutional theories to motivate their empirical investigations.

A caveat is in order, though: the forces driving location and relocation processes differ. This is not always apparent because location theories tend to overemphasise (minimise) the importance of “pull” (“push”) factors in relocation decisions. However, relocations result from a sequential decision conditional upon the history of the firm or the establishment. That is, relocation decisions are taken conditional upon previous location decisions (Pellenbarg et al. 2002b). It is therefore plausible to conclude that the information used to take the decision of (re)locating here or there is not the same. In particular, migrations within the same geographical market are likely to have more and better information about the sites than start-ups.

Notice, however, that this theoretical framework says little about the sense of these differences in the determinants of industrial (re)location. This means that a priori we cannot predict which neoclassical, behavioural and institutional factors will (not) affect location and relocation decisions. It is also not clear whether it is different factors that affect locations and relocations or it is the same factors that affect both processes but with different intensity. Consequently, it seems that a sensible empirical strategy is to “let the data speak” and use the same vector of explanatory variables for the rate of both start-ups and relocations.

### **3.2 Empirical studies**

Recent surveys by Arauzo et al. (2006), Basile and McCann (2006) and Blonigen (2006) provide an excellent overview of the empirical literature on industrial location. However, this evidence basically refers to new concerns. Inferences from relocation data are much less common, especially those obtained

conditioning on a set of explanatory variables. Pellenbarg et al. (2002a, 2002b) review the firm relocation literature since WWII until the late 1990s and find indeed that descriptive statistical methods prevail. Interesting as this research might be, it says very little about casual effects.

As for the studies providing sounded econometric evidence, we have already mentioned the paper of Holl (2004a) because she makes use of CDM. However, there is a number of investigations on relocation that do not use CDM but DCM. To the best of our knowledge, these investigations include Erickson and Wasylenko (1980), Baudewyns et al. (2000), van Dijk and Pellenbarg (2000), Brower et al. (2004), Lee (2004) and Strauss-Kahn and Vives (2005). Next we shall briefly review these papers in terms of data, aims and econometric specification.

Three different sources of data have been used in this literature. Baudewyns et al. (2000), van Dijk and Pellenbarg (2000), Brower et al. (2004) and Strauss-Kahn and Vives (2005) exploit survey data, whereas Erickson and Wasylenko (1980) focus on a sample of firms that moved from Milwaukee City to its suburbs in the 1964 to 1974 period and Lee (2004) analyses census data (as we do). The statistical implications of using these distinct sources are worth noting. First, there is the issue of sample representativeness, apparent for example in Strauss-Kahn and Vives (2005) and Erickson and Wasylenko (1980). Inferences are still valid conditioning on the sample, though authors do not always make this point clear enough –see e.g. van Dijk and Pellenbarg (2000) and Brower et al. (2004). Second, as Barkley and McNamara (1994) show, stated preferences may be a poor proxy for location decisions (revealed preferences). Inferences from survey data and ad-hoc defined samples should therefore be interpreted with care.

These caveats aside, the aim of Baudewyns et al. (2000) is to analyse the effects that better public infrastructures had on the (de)location decisions of Belgium firms from the city of Brussels (in the period 1981 to 1991) and the region of Wallonia (in the period 1990 to 1994) using a conditional logit model. Agglomeration economies and wage levels were also included among the regressors. Data come from STRATED for Brussels and Dun & Bradstreet for Wallonia. As for Van Dijk and Pellenbarg (2000),



they analyse the stated preference of Dutch firms with regard to migration (measured as the propensity to move within the next two years reported in a questionnaire launched in 1995-1996) using ordered logit and probit models. Thanks to the detailed information they collected from the questionnaire, these authors are able to distinguish between internal (organisational structure, financial reserves, size, etc.), external (labour market characteristics, government policies, general economic conditions, etc.) and location (occupancy characteristics, accessibility, distance to suppliers and markets, etc.) factors as determinants of relocation decisions.

Whereas these two papers analyse firms from the same country, Brower et al. (2004) and Strauss-Kahn and Vives (2005) analyse firms from different countries. In particular, Brower et al. (2004) analyse a sample of firms with more than 200 employees and from twenty-one countries. The dependent variable is binary ("In the 1999 [Cranet] survey, respondents were asked whether their firms have relocated in the last three years"), the econometric specification is the logit model and the main covariates are age, size, sector, size of the market and the region where the firm is located. As for Strauss-Kahn and Vives (2005), they use a nested logit to discern between the decision of "where to locate" and that of "whether relocate". The explanatory variables are wages, population, the sectorial distribution of employment, measures of agglomeration economies, corporate taxes and accessibility to airports. Data come from Dun & Bradstreet and official statistics. Also, two distinct features are worth mentioning about this paper: it presents a supportive theoretical model and the unit of analysis are US headquarters (from both US and non-US companies) rather than firms or plants.

The aim of Ericsson and Wasylenko (1980) is to study intra-metropolitan firm relocation. They first derive a demand function for land from two simple models of cost minimisation and profit maximisation strategies. They use a logistic model for the decision of leaving the central city (Milwaukee) and agglomeration economies, accessibility to labour supply and available land as explanatory variables. The sample includes firms from manufacturing and non-manufacturing industries. Data comes from unemployment compensation records compiled by the Wisconsin Department of Industry, Labor and Human Relations. This paper also presents a supportive theoretical model

Lastly, Lee (2004) aims to assess the impact of state development incentives on the decisions to (re)locate. He uses a multinomial logit model in which the dependent variable can take three values: entry, relocation and exit. Rather than analysing the parameter estimates, however, in a second stage he evaluates the predicted probabilities for shutting down and relocating in a scenario of change vs. maintenance of the incentive policy. “The difference in those two sets of adjusted probabilities is the difference due to the policy change, holding other characteristics constant”. Moreover, he uses the same “method of recycled predictions” in a probit model for the decision to relocate within or away from the state. The explanatory variables consist of a vector of dummies for each incentive program as well as firm- and location-specific variables. Data covers the universe of US manufacturing establishments between 1972 and 1992, though he focuses in multi-plant firms that relocated production (defined as the opening of a new plant more than 50 miles from the county of the original location, producing in the same four-digit industry and implying a reduction of more than 50% of the total employment in the original location).

## **4. Data and results**

### **4.1 The data-base**

Table 1 reports details on the definition, source and descriptive statistics of the dependent and explanatory variables used in this study. The data-base covers 946 municipalities in the period 2001 to 2004. However, Table 1 shows that there are some (random) missing data in the explanatory variables that make our final sample an unbalanced panel. We found no clear pattern in these missing values, i.e. they are not concentrated in e.g. small villages or in a particular province. It was consequently judged unnecessary to implement corrections for sample selection bias.

[Insert Table 1 about here]

The dependent variables in Panel A of Table 1 were calculated from the information available in the “Registry of Industrial Establishments of Catalonia”. This administrative register contains all the

establishments located in Catalan municipalities. However, since 2001 it provides (upon request) separated data for new and relocated establishments. To make the sample sectorially homogeneous we selected all establishments in codes 12 to 36 of the NACE-93 classification of the European Union. We have therefore limited the study to industrial establishments.

Information regarding the explanatory variables comes basically from the data-base constructed by Trullén and Boix (2004). This was complemented with data from the Catalan Institute of Statistics. As Panel B of Table 1 shows, we have no variables related to the behavioural location theory and dummies concerning the spatial organization of Catalonia are the only measures related to the institutional location theory. In other words, most regressors refer to the neoclassical location theory.

This specification may raise concerns about the omission of relevant variables. However, it is very difficult to assess the statistical relevance of these concerns in non-linear models like the ones used in this study. The fact that no empirical evidence exists showing the relative importance of each theory in location decisions does not help either. In any case, given that our unit of analysis is the municipality and not the firm the potential biases caused by the lack of behavioural factors should be small. As Pellenbarg et al. (2002a, 2002b) argue, behavioural theory is mostly a theory of entrepreneurship, whereas neoclassical and institutional theories appear to respectively explain better the behaviour of small and large firms. We have no information on the number of entrepreneurs in our data-set, but we do know the size of the establishments: four out of five have less than 10 employees and one per cent of the establishments in the sample have more than 50 employees. It seems therefore that there is little risk of misspecification in our vector of explanatory variables.

The lack of behavioural factors contrasts however with the richness of information on certain institutional (metropolitan areas) and neoclassical (location economies, human capital and transport infrastructure) factors. To facilitate the interpretation of results, in Tables 2 to 5 we report Wald tests for these sets of coefficients rather than the estimated coefficients of each variable. “Wald Test L.E.”, for example, stands for the Wald test for the joint significance of the coefficients associated with the

seventeen measures of location economies. Similarly for “Wald Test H.C.” (human capital), “Wald Test Infra.” (transport infrastructures) and “Wald Test Met.” (metropolitan areas).

It is also worth mentioning that we have two measures of urbanization and location economies available. One is based on the total area of the municipality and the other on the area of the municipality defined as urban land. In general, results did not change substantially when using one or the other. The exception being the coefficients of density and (dis)urbanisation economies. These were not statistically significant when using measures based on the area of the municipality. We consequently opted for reporting those based on urban land, which seems a more accurate measure and provides indeed more consistent results.

## 4.2 Estimates

In this section we present results from the estimation of count data models for the expected number of new and relocated industrial establishments created in Catalonia in the period 2001 to 2004, conditional to the same set of explanatory factors (summarised in Panel B of Table 1). In Tables 2 to 4 we present results under alternative assumptions on the exposure period and independence of the events, whereas in Table 5 we address the independence assumption. Besides a number of statistics (Wald tests, log L, etc.), we report partial or marginal effects. We do not report coefficient estimates because our interest here is to empirically determine whether the opening of new and relocated establishments are processes driven by different factors or by the same factors but with different intensity. We are therefore interested in the “average response” of the dependent variables, which does not correspond to coefficient estimates in non-linear models like the ones used here.<sup>8</sup>

Before presenting the econometric evidence, however, it seems interesting to test for the equality of means of the dependent variables. After all, there is little point in running separate regressions for start-ups and relocations if we cannot reject the null hypothesis of this test. To this end we use a classical

---

<sup>8</sup> For covariates that are continuous, changes in the conditional mean of start-ups and relocations are computed at the mean of the covariates; for dummies, it is the difference in the prediction of the dependent variable associated with the 0-to-1 change in the covariate. See Cameron and Trivedi (1998, 2001) for details on the way to do these calculations.

conditional test and an unconditional test recently proposed by Krishnamoorthy and Thomson (2004). We test for the equality of means using yearly and pooled data. All the performed tests rejected the null hypothesis. We can now confidently turn to analyse the inferences obtained from our models.

#### **4. Conclusions**

This paper analyses the determinants of industrial location distinguishing between start-ups and relocations. We discuss several econometric strategies based upon count regression models for cross-section and panel data, thus assessing the robustness of our conclusions to alternative distributional assumptions about the data generating process. Interestingly, in this framework we cannot only test for the distinct characteristics that make a municipality more attractive to start-ups and relocations but also for the existence of a relation between these events.

We provide evidence that the location of new industrial establishments and the relocation of extant industrial establishments are driven by similar stochastic processes and determinants. However, the weight of these determinants, measured by the partial or marginal effect, differs between start-ups and relocations. Also, we find that locations and relocations are positively interrelated, which can be interpreted as evidence of some kind of agglomeration economies.

From an economic policy viewpoint, results indicate that public programs aiming to attract new businesses should pay attention to these issues. Otherwise, they are likely to fail in their objectives. However, the proposed econometric specifications enable us to address specific questions on the factors associated with the public sector (infrastructures, human capital, etc.), the private sector (e.g. agglomeration and location economies) and/or the space (distance to capitals, territorial dummies, etc.) that affect the expected number of new and relocated establishments per municipality.

## References

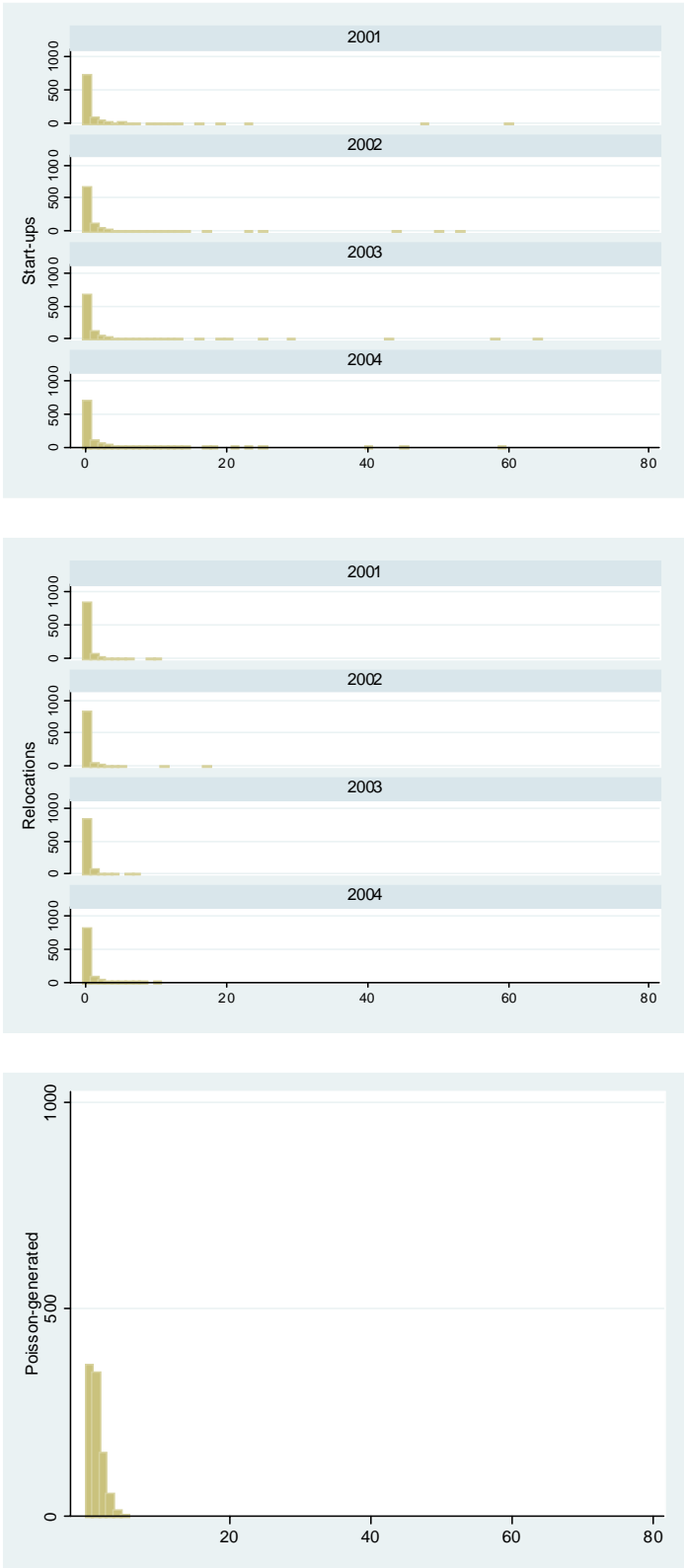
- Arauzo, J.M. (2005): “Determinants of Industrial Location. An Application for Catalan Municipalities”, *Papers in Regional Science* **84** (1): 105-120.
- Arauzo, J.M. and Manjón, M. (2004): “Firm Size and Geographical Aggregation: An Empirical Appraisal in Industrial Location”, *Small Business Economics* **22**: 299-312.
- Arauzo, J.M., Liviano, D. and Manjón, M. (2006): “Empirical Studies in Industrial Location: An Assessment of Their Methods and Results”, mimeo.
- Barkley, D.L. and McNamara, K.T. (1994): “Manufacturer’s Location Decisions: Do Surveys Provide Helpful Insights?” *International Regional Science Review* **17**(1): 23-47.
- Basile, R. and McCann, P. (2006): “Location Behaviour of Multinational Firms: a Review of the Literature”, mimeo.
- Baudewyns, D.; Ben Ayad, M. and Sekkat, K. (2000): “Infrastructure publique et localisation des entreprises à Bruxelles et en Wallonie”, in M. Beine and F. Docquier (eds.), *La politique de développement local et l'infrastructure publique: Bruxelles et Wallonie*, Bruxelles.
- Becker, R. and Henderson, V. (2000): “Effects of Air Quality Regulations on Polluting Industries”, *Journal of Political Economy* **108** (2): 379– 421.
- Blonigen, B.A. (2006): “A Review of the Empirical Literature on FDI Determinants,” *Atlantic Economic Journal*, forthcoming.
- Brouwer, A.E., Mariotti, I. and van Ommeren, J.N. (2004): “The Firm Relocation Decision: An Empirical Investigation”, *The Annals of Regional Science* **38**: 335– 347.
- Cameron, A.C. and Trivedi, P.K (1998): *Regression Analysis of Count Data*, Cambridge University Press, Cambridge.
- Cameron, A.C. and Trivedi, P.K (2001): “Essentials of Count Data Regression”, in B. H. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Blackwell, Oxford.
- Carlton, D. (1979): “Why New Firms Locate Where They Do: An Econometric Model”, in W. Wheaton (ed.), *Interregional Movements and Regional Growth*. The Urban Institute, Washington.
- van Dijk, J. Pellenbarg, P. H. (2000): “Firm Relocation Decisions in the Netherlands: An Ordered Logit Approach”, *Papers in Regional Science* **79**: 191-219.

- Erickson, R.A. and Wasylenko, M. (1980): "Firm Relocation and Site Selection in Suburban Municipalities", *Journal of Urban Economics* **8**: 69-85.
- Figueiredo, O.; Guimarães, P. and Woodward, D. (2002): "Home-field Advantage: Location Decisions of Portuguese Entrepreneurs", *Journal of Urban Economics* **52**: 341-361.
- Frenkel, A. (2001): "Why High-Technology Firms Choose to Locate in or near Metropolitan Areas", *Urban Studies* **38**(7): 1083-1101.
- Guimarães, P.; Figueiredo, O. and Woodward, D. (2003): "A Tractable Approach to the Firm Location Decision Problem", *Review of Economics and Statistics* **85**(1): 201-204.
- Guimarães, P.; Figueiredo, O. and Woodward, D. (2004): "Industrial Location Modeling: Extending the Random Utility Framework", *Journal of Regional Science* **44** (1): 1-20.
- Hausman, J.A.; Hall, B.A. and Griliches, Z. (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship", *Econometrica* **52** (4): 909-938.
- Hayter, R. (1997): *The dynamics of industrial location. The factory, the firm and the production system*, New York: Wiley.
- Holl, A. (2004a): "Start-ups and Relocations: Manufacturing Plant Location in Portugal", *Papers in Regional Science* **83** (4): 649-668.
- Holl, A. (2004b): "Transport Infrastructure, Agglomeration Economies, and Firm Birth. Empirical Evidence from Portugal", *Journal of Regional Science* **44** (4): 693-712.
- Holl, A. (2004c): "Manufacturing Location and Impacts of Road Transport Infrastructure: Empirical Evidence from Spain", *Regional Science and Urban Economics* **34** (3): 341-363.
- Krishnamoorthy, K. and Thomson, J. (2004): "A more powerful test for comparing two Poisson means", *Journal of Statistical Planning and Inference* **119**: 23-35.
- Lee, Y. (2004): "Geographic Redistribution of US Manufacturing and the Role of State Development Policy", Working Paper 04-15, Federal Reserve Bank of Cleveland.
- List, J.A. and McHone, W.W. (2000): "Measuring the Effects of Air Quality Regulations on "Dirty" Firm Births: Evidence from the Neo and Mature-Regulatory Periods", *Papers in Regional Science* **79**: 177-190.

- Mullahy, J. (1997): “Heterogeneity, Excess Zeros, and the Structure of Count Data Models”, *Journal of Applied Econometrics* **12** (3): 337-350.
- Papke, L. (1991): “Interstate Business Tax Differentials and New Firm Location”, *Journal of Public Economics* **45**: 47-68.
- Pellenbarg, P.H., van Wissen, L.J.G. and van Dijk, J. (2002a): “Firm Relocation: State of the Art and Research Prospects”, SOM Research Report 02D31, University of Groningen.
- Pellenbarg, P.H., van Wissen, L.J.G. and van Dijk, J. (2002b): “Firm Migration”, in P. McCann (ed.), *Industrial location Economics*. Edward Elgar Publishing, Cheltenham.
- Strauss-Kahn, V. and Vives, X. (2005): “Why and Where Do Headquarters Move?”, Discussion Paper No. 5070, Centre for Economic Policy Research.
- Trullén, J. and Boix, R. (2004): *Indicadors 2005*, Diputació de Barcelona i Universitat Autònoma de Barcelona.



**Figure 1:** Frequency distribution of Start-ups, Relocations and Poisson-Distributed Random Variable with Unitary Mean.



**Table 1: Descriptive statistics (pooled sample).**

	Period	Obs.	Mean	S. D.	Min.	Max.
<b>Panel A: Dependent Variables</b>						
Start-ups	2001-2004	3784	0.93	3.55	0	64
Relocations	2001-2004	3784	0.19	0.75	0	17
<b>Panel B: Explanatory Variables</b>						
<b>B.1 Neoclassical approach</b>						
Population density ( <i>DENS</i> )	2001-2004	3530	40.43	100.13	0.74	3220
Urbanisation economies: Workers per km <sup>2</sup> ( <i>URB</i> )	2001-2004	3530	10.47	33.73	0	980
Location Economies: Workers in sector “i” ( $i = 1, \dots, 17$ ) per km <sup>2</sup> .	2001-2004	3648	-	-	-	-
Industrial diversity: H-H index of diversity of industrial jobs ( <i>DIV</i> )	2001-2004	3758	3.57	1.41	1	7.43
Industry share: % of jobs in the industrial sector ( <i>INDS</i> )	2001	3784	0.22	0.11	0	0.60
Services share: % of jobs in the services sector ( <i>SERS</i> )	2001-2004	3758	0.48	0.24	0	1
Entrepreneurship: % of jobs declared to be entrepreneurs ( <i>ENTREP</i> )	2001	3784	0.26	0.11	0.02	0.69
Commuting: Mobility ratio ( <i>COMM</i> )	2001	3780	1.79	17.30	0	521.07
Human capital: % of population working in science and technology	2001	3784	0.19	0.05	0.06	0.47
Human capital: % of population with a university degree	2001	3784	0.14	0.05	0.01	0.5
Human capital: Years of education, average for population over 25 years old	2001	3784	8.5	1.01	4.23	11.99
Infrastructures: Average travelling time to capital of province	2001	3768	87.38	23.30	56	190
Infrastructures: Average travelling time to nearest airport	2001	3768	49.07	32.98	0	190
Infrastructures: Average travelling time to nearest port	2001	3768	62.44	32.99	0	197
Infrastructures: Dummy for train station	2001-2004	3780	0.10	0.30	0	1
<b>B.2 Institutional approach</b>						
Dummy for the province of Barcelona ( <i>PROV1</i> )	2001-2004	3784	0.32	0.46	0	1
Dummy for the province of Girona ( <i>PROV2</i> )	2001-2004	3784	0.23	0.42	0	1
Dummy for the province of Lleida ( <i>PROV3</i> )	2001-2004	3784	0.24	0.42	0	1
Dummy for capital of “comarca” ( <i>CAPC</i> )	2001-2004	3784	0.04	0.20	0	1
Dummy for coastal municipality ( <i>COAST</i> )	2001-2004	3784	0.07	0.26	0	1
Dummy for metropolitan area of Barcelona	2001-2004	3784	0.23	0.42	0	1
Dummy for metropolitan area of Girona	2001-2004	3784	0.07	0.26	0	1
Dummy for metropolitan area of Lleida	2001-2004	3784	0.06	0.24	0	1
Dummy for metropolitan area of Manresa	2001-2004	3784	0.03	0.17	0	1
Dummy for metropolitan area of Tarragona	2001-2004	3784	0.07	0.26	0	1

**Source:** “Registry of Industrial Establishments of Catalonia” (dependent variables), Trullen and Boix (2004) and Catalan Institute of Statistics (explanatory variables).

**Note:** Details on the measures of location economies are not reported to save space. The fourth province of Catalonia (not reported) is Tarragona. A “comarca” is a territorial unit formed by adjacent municipalities. There are 11 “comarques” in the province of Barcelona, 8 in Girona, 12 in Lleida and 10 in Tarragona.

**Table 2: Determinants of Start-ups and Relocations (cumulative and period means data).**

	Start-ups			Relocations		
	NBM	ZIPM	NBIM	NBM	ZIPM	NBIM
<i>DENS</i>	-0.0134 (0.0026)***	-0.0130 (0.0016)***	-0.0162 (0.0035)***	-0.0031 (0.0007)***	-0.0045 (0.0012)***	-0.0059 (0.0016)***
<i>URB</i>	0.0358 (0.0080)***	0.0655 (0.0054)***	0.0536 (0.0120)***	0.0098 (0.0022)***	0.0232 (0.0042)***	0.0229 (0.0055)***
<i>URBA</i> <sup>2</sup>	-7.2×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )***	-0.0001 (1.0×10 <sup>-5</sup> )***	-9.9×10 <sup>-5</sup> (-2.0×10 <sup>-5</sup> )***	-1.6×10 <sup>-5</sup> (-8.8×10 <sup>-5</sup> )***	-4.3×10 <sup>-5</sup> (-1.0×10 <sup>-5</sup> )***	-4.6×10 <sup>-5</sup> (-1.0×10 <sup>-5</sup> )***
<i>DIV</i>	0.3568 (0.0386)***	0.1939 (0.0298)***	0.3650 (0.0513)***	0.0810 (0.0123)***	0.0800 (0.0211)***	0.1016 (0.0243)***
<i>INDS</i>	0.9059 (0.6404)	3.3522 (0.6260)***	1.5654 (0.9699)*	0.3847 (0.2037)*	1.5431 (0.4620)***	1.0238 (0.4870)**
<i>SERS</i>	-0.3773 (0.2607)	1.0607 (0.2718)***	-0.1377 (0.4073)	-0.0364 (0.0918)	0.0298 (0.1923)	0.0280 (0.2133)
<i>ENTREP</i>	-3.6275 (0.7654)***	-5.6298 (0.8339)***	-3.4208 (1.1508)***	-0.8115 (0.2588)***	-1.3407 (0.5794)**	-1.0641 (0.5984)*
<i>COMM</i>	0.0011 (0.0014)	-3.7×10 <sup>-5</sup> (0.0006)	0.0013 (0.0018)	0.0001 (0.0003)	0.0004 (0.0004)	0.0002 (0.0006)
<i>PROV1</i>	0.8743 (0.2362)***	1.6899 (0.2589)***	1.2942 (0.3419)***	0.2566 (0.1011)**	0.6602 (0.2213)***	0.6687 (0.2523)***
<i>PROV2</i>	0.6197 (0.3526)*	1.1193 (0.3422)***	1.0036 (0.5271)*	0.3859 (0.1778)**	0.9373 (0.3595)***	0.8496 (0.3895)**
<i>PROV3</i>	1.2683 (0.4502)***	2.5751 (0.4996)***	1.8946 (0.6782)***	0.4107 (0.2051)**	0.8809 (0.3976)**	0.9258 (0.4718)**
<i>CAPCOM</i>	0.2675 (0.2580)	1.2935 (0.2327)***	0.5235 (0.3834)	0.2195 (0.1155)*	0.37701 (0.1699)**	0.4201 (0.2278)*
<i>COAST</i>	0.8961 (0.2896)***	0.8067 (0.1427)***	0.8834 (0.3268)***	0.1884 (0.0801)**	0.4020 (0.1239)***	0.2563 (0.1332)*
Wald Test L.E.	62.17***	472.30***	61.91***	63.36***	86.47***	61.18***
Wald Test H.C.	8.01**	31.66***	6.83*	4.54	6.71*	5.47
Wald Test Infra.	19.32***	64.74***	17.01***	20.20***	24.74***	19.64***
Wald Test Met.	2.97	76.32***	4.79	13.12**	17.02***	8.15
Log L	-1357.13	-1743.93	-1330.92	-680.23	-682.52	-650.43
LR Joint Test	897.08***	5663.09***	648.78***	570.79***	667.71***	292.43***
GoF Test						
LR $\alpha$ Test	1119.41***			121.22***		
Vuong Test		3.67***	3.20***		4.05***	3.58***
LR ZIPM-NBIM Test		826.03***			64.18***	

Note: Marginal or partial effects of selected explanatory variables (see Panel B of Table 1) are reported. Standard errors in brackets. The symbols \*\*\*, \*\* and \* denote significance at 1%, 5% and 10% levels, respectively (907 observations). See sections 2 and 4 for definitions of the reported statistics.

**Table 3.A: Determinants of Start-ups and Relocations (2001 data).**

	Start-ups			Relocations		
	NBM	ZIPM	NBIM	NBM	ZIPM	NBIM
<i>DENS</i>	-0.0015 (0.0004)***	-0.0024 (0.0007)***	-0.0028 (0.0010)***	-0.0002 (0.0001)*	-0.0004 (0.0002)*	-0.0005 (0.0003)
<i>URB</i>	0.0078 (0.0019)***	0.0204 (0.0046)***	0.0193 (0.0054)***	0.0010 (0.0004)**	0.0025 (0.0011)**	0.0027 (0.0013)**
<i>URBA</i> <sup>2</sup>	-2.2×10 <sup>-5</sup> (0.0001)***	-4.8×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )***	-5.6×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )***	-6.3×10 <sup>-6</sup> (1.0×10 <sup>-5</sup> )***	-1.5×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )**	-1.5×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )**
<i>DIV</i>	0.0355 (0.0084)***	0.0237 (0.0131)*	0.0475 (0.0165)***	0.0046 (0.0025)*	0.0060 (0.0038)	0.0074 (0.0048)
<i>INDS</i>	0.1887 (0.1266)	1.1434 (0.3255)***	0.5307 (0.3527)	0.0411 (0.0310)	0.1489 (0.0949)	0.1444 (0.1037)
<i>SERS</i>	-0.0494 (0.0540)	0.0825 (0.1285)	-0.0018 (0.1517)	0.0176 (0.0133)	0.0787 (0.0468)*	0.0805 (0.0530)
<i>ENTREP</i>	-0.4438 (0.1954)**	-0.8330 (0.4257)**	-0.5005 (0.4982)	-0.0071 (0.0344)	0.1212 (0.1107)	0.1214 (0.1203)
<i>COMM</i>	-0.0008 (0.0014)	-0.0023 (0.0024)	-0.0021 (0.0032)	0.0002 (0.0002)	0.0005 (0.0005)	0.0005 (0.0006)
<i>PROV1</i>	-0.0077 (0.0382)	0.1213 (0.0932)	0.0287 (0.0984)	0.01450 (0.0149)	0.0463 (0.0424)	0.0494 (0.0486)
<i>PROV2</i>	-0.0283 (0.0379)	-0.0856 (0.0800)	-0.0422 (0.0991)	0.0129 (0.0173)	0.0392 (0.0489)	0.0433 (0.0549)
<i>PROV3</i>	0.0144 (0.0576)	0.1931 (0.1606)	0.0231 (0.1371)	0.0193 (0.0240)	0.0721 (0.0785)	0.0833 (0.0950)
<i>CAPCOM</i>	0.0799 (0.0597)	0.3470 (0.1352)***	0.2117 (0.1459)	0.0314 (0.0275)	0.0681 (0.0571)	0.0800 (0.0718)
<i>COAST</i>	0.0762 (0.0426)*	0.2497 (0.0849)***	0.1334 (0.0853)	0.0022 (0.0062)	0.0150 (0.0169)	0.0051 (0.0162)
Wald Test L.E.	38.94***	123.26***	37.83***	18.96	29.54**	22.02
Wald Test H.C.	7.35*	14.02***	8.42*	10.01**	10.46**	8.79**
Wald Test Infra.	3.92	10.17*	1.87	4.52	5.17	5.13
Wald Test Met.	4.09	16.64***	4.64	10.49**	8.52	8.02
Log L	-596.51	-612.07	-576.52	-237.77	-228.19	-227.24
LR Joint Test	522.26***	873.75***	266.93***	264.13***	180.02***	131.42***
GoF Test						
LR $\alpha$ Test	124.28***			4.95***		
Vuong Test		3.56***	3.33***		2.07**	1.96**
LR ZIPM-NBIM Test		71.09***				1.88*

Note: Marginal or partial effects of selected explanatory variables (see Panel B of Table 1) are reported. Standard errors in brackets. The symbols \*\*\*, \*\* and \* denote significance at 1%, 5% and 10% levels, respectively (816 observations). See sections 2 and 4 for definitions of the reported statistics.

**Table 3.B: Determinants of Start-ups and Relocations (pooled data).**

	Start-ups			Relocations		
	NBM	ZIPM	NBIM	NBM	ZIPM	NBIM
<i>DENS</i>	-0.0028 (0.0004)***	-0.0036 (0.0005)***	-0.0045 (0.0008)***	-0.0005 (0.0001)***	-0.0011 (0.0003)***	-0.0012 (0.0003)***
<i>URB</i>	0.0097 (0.0013)***	0.0173 (0.0017)***	0.0166 (0.0025)***	0.0024 (0.0004)***	0.0061 (0.0011)***	0.0060 (0.0013)***
<i>URBA</i> <sup>2</sup>	-1.4×10 <sup>-5</sup> (0.0001)***	-2.6×10 <sup>-5</sup> (0.0002)***	-2.8×10 <sup>-5</sup> (0.0001)***	-2.8×10 <sup>-6</sup> (1.1×10 <sup>-5</sup> )***	-2.8×10 <sup>-5</sup> (3.8×10 <sup>-5</sup> )***	-2.8×10 <sup>-5</sup> (3.8×10 <sup>-5</sup> )***
<i>DIV</i>	0.0827 (0.0072)***	0.0437 (0.0092)***	0.0894 (0.0122)***	0.0194 (0.0028)***	0.0209 (0.0059)***	0.0250 (0.0056)***
<i>INDS</i>	0.2313 (0.1326)*	1.0612 (0.2018)***	0.5444 (0.2530)**	0.0991 (0.0494)**	0.4186 (0.1354)***	0.2613 (0.1205)**
<i>SERS</i>	-0.0331 (0.0536)	0.1847 (0.0787)**	0.0552 (0.1066)	0.0087 (0.0209)	0.0169 (0.0542)	0.0278 (0.0508)
<i>ENTREP</i>	-1.1805 (0.1658)***	-1.7199 (0.2542)***	-1.2165 (0.3137)***	-0.2728 (0.0634)***	-0.2656 (0.1688)	-0.2305 (0.1512)
<i>COMM</i>	0.0002 (0.0002)	-0.0001 (0.0002)	0.0002 (0.0004)	6.5×10 <sup>-5</sup> (8.0×10 <sup>-5</sup> )	0.0001 (0.0001)	0.0001 (0.0001)
<i>PROV1</i>	0.2707 (0.0538)***	0.5216 (0.0892)***	0.4862 (0.0988)***	0.0721 (0.0243)***	0.1830 (0.0625)***	0.1759 (0.0614)***
<i>PROV2</i>	0.2214 (0.0794)***	0.3949 (0.1191)***	0.3683 (0.1374)***	0.0981 (0.0394)**	0.2383 (0.0950)**	0.2205 (0.0913)**
<i>PROV3</i>	0.3774 (0.0996)***	0.8488 (0.1731)***	0.6360 (0.1772)***	0.1064 (0.0479)**	0.2488 (0.1130)**	0.2439 (0.1133)**
<i>CAPCOM</i>	0.1812 (0.0634)***	0.3891 (0.0738)***	0.2770 (0.1022)***	0.0817 (0.0295)***	0.1212 (0.0508)**	0.1261 (0.0549)**
<i>COAST</i>	0.1804 (0.0475)***	0.2500 (0.0447)***	0.2019 (0.0671)***	0.0464 (0.0169)***	0.1127 (0.0357)***	0.0827 (0.0335)**
Wald Test L.E.	100.12***	449.60***	113.59***	73.84***	86.28***	71.92***
Wald Test H.C.	16.02***	29.89***	12.84***	8.43**	7.41*	6.99*
Wald Test Infra.	36.95***	59.59***	26.82***	27.16***	23.60***	22.94***
Wald Test Met.	6.30	72.27***	8.93	17.09***	18.62***	11.27**
Log L	-2919.64	-3124.23	-2850.60	-1283.27	-1272.44	-1243.77
LR Joint Test	2107.52***	3945.33***	1172.80***	982.15***	541.59***	408.94***
GoF Test						
LR $\alpha$ Test	844.60***			69.34***		
Vuong Test		6.87***	4.92***		3.66***	4.03***
LR ZIPM-NBIM Test			547.25***			57.34***

Note: Marginal or partial effects of selected explanatory variables (see Panel B of Table 1) are reported. Standard errors in brackets. The symbols \*\*\*, \*\* and \* denote significance at 1%, 5% and 10% levels, respectively (3499 observations). See sections 2 and 4 for definitions of the reported statistics.

**Table 4: Determinants of Start-ups and Relocations (panel data).**

	Start-ups				Relocations			
	POIS-FE	POIS-RE	NB-FE	NB-RE	POIS-FE	POIS-RE	NB-FE	NB-RE
<i>DENS</i>	0.0126 (0.0070)*	-0.0081 (0.0027)***	-0.0180 (0.0069)***	-0.0080 (0.0026)***	0.0009 (0.0111)	-0.0117 (0.0035)***	-0.0093 (0.0102)	-0.0112 (0.0039)***
<i>URB</i>	-0.0072 (0.0093)	0.0314 (0.0069)***	-0.0046 (0.0111)	0.0309 (0.0070)***	-0.0149 (0.0238)	0.0454 (0.0106)***	0.0160 (0.0267)	0.0374 (0.0114)***
<i>URBA</i> <sup>2</sup>	1.4×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )	-3.8×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )***	1.7×10 <sup>-5</sup> (2.0×10 <sup>-5</sup> )	-3.7×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )***	1.5×10 <sup>-5</sup> (2.0×10 <sup>-5</sup> )	-4.9×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )***	-9.9×10 <sup>-5</sup> (8.0×10 <sup>-5</sup> )	-3.8×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )***
<i>DIV</i>	0.0606 (0.0712)	0.3204 (0.0374)***	0.0967 (0.0752)	0.3369 (0.0373)***	0.1537 (0.1602)	0.4006 (0.0571)***	0.0847 (0.1463)	0.3707 (0.0607)***
<i>INDS</i>	-	0.7424 (0.7668)	-	0.4890 (0.7615)	-	1.7982 (1.0894)*	-	1.2159 (1.1659)
<i>SERS</i>	-0.3202 (0.6691)	-0.1635 (0.2754)	-0.3461 (0.6792)	-0.1884 (0.2800)	2.5295 (1.3975)*	0.2647 (0.4461)	-0.9719 (1.1726)	0.2566 (0.4676)
<i>ENTREP</i>	-	-5.3736 (0.9472)***	-	-5.4259 (0.9678)***	-	-4.7148 (1.4274)***	-	-5.3543 (1.5052)***
<i>COMM</i>	-	0.0014 (0.0018)	-	0.0014 (0.0017)	-	0.0013 (0.0019)	-	0.0013 (0.0021)
<i>PROV1</i>	-	1.0417 (0.2033)***	-	1.1046 (0.2077)***	-	1.0945 (0.3411)***	-	0.9934 (0.3502)***
<i>PROV2</i>	-	0.7248 (0.3002)**	-	0.7029 (0.3042)**	-	1.3695 (0.4114)***	-	1.2884 (0.4295)***
<i>PROV3</i>	-	1.2970 (0.2900)***	-	1.2322 (0.2889)***	-	1.3899 (0.4665)***	-	1.2357 (0.4819)***
<i>CAPCOM</i>	-	0.7745 (0.2412)***	-	0.7526 (0.2325)***	-	1.0369 (0.2949)***	-	1.1271 (0.3215)***
<i>COAST</i>	-	0.6307 (0.1794)***	-	0.6383 (0.1796)***	-	0.6602 (0.2290)***	-	0.4914 (0.2574)*
Wald Test L.E.	17.25	29.31**	40.02***	35.45***	4.76	49.12***	13.88	35.97***
Wald Test H.C.	-	11.42***	-	13.29***	-	7.40*	-	4.89
Wald Test Infra.	-	23.81***	-	19.13***	-	22.34***	-	18.16***
Wald Test Met.	-	2.73	-	2.15	-	13.96**	-	10.72*
Log L	-1400.71	-2788.12	-1385.40	-2760.66	-568.66	-1261.65	-611.42	-1286.03
LR Joint Test	18.68	1103.82***	46.88***	1121.63***	10.15	568.53***	69.56***	450.77***
Hausman Test	66.63***		62.94***		15.80		26.81	

Note: POIS(NB)-FE and POIS(NB)-RE denote Poisson (Negative Binomial) Fixed and Random Effects, respectively. Marginal or partial effects of selected explanatory variables (see Panel B of Table 1) are reported. Standard errors in brackets. The symbols \*\*\*, \*\*, and \* denote significance at 1%, 5% and 10% levels, respectively (1634 and 3499 observations for FE and RE estimates, respectively). See sections 2 and 4 for definitions of the reported statistics. Estimates from POIS-FE, NB-FE (Start-ups) and NB-RE (Relocations) were obtained under weak convergence criteria.

**Table 5: Determinants of Start-ups and Relocations (cumulative, period means data and interrelations).**

	Start-ups			Relocations		
	NBM	ZIPM	NBIM	NBM	ZIPM	NBIM
<i>DENS</i>	-0.0084 (0.0020)***	-0.0071 (0.0017)***	-0.0099 (0.0030)***	-0.0020 (0.0006)***	-0.0045 (0.0014)***	-0.0047 (0.0015)***
<i>URB</i>	0.0290 (0.0062)***	0.0570 (0.0060)***	0.0467 (0.0095)***	0.0075 (0.0019)***	0.0202 (0.0044)***	0.0181 (0.0051)***
<i>URBA</i> <sup>2</sup>	-5.7×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )***	-9.1×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )***	-7.9×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )***	-1.4×10 <sup>-5</sup> (7.0×10 <sup>-5</sup> )***	-4.8×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )***	-4.5×10 <sup>-5</sup> (1.0×10 <sup>-5</sup> )***
<i>DIV</i>	0.2719 (0.0316)***	0.1968 (0.0335)***	0.2753 (0.0472)***	0.0636 (0.0108)***	0.0954 (0.0238)***	0.0911 (0.0246)***
<i>INDS</i>	0.4832 (0.5515)	0.0299 (0.7114)	0.6782 (0.8982)	0.3393 (0.1877)*	0.7268 (0.4672)	0.7413 (0.4830)
<i>SERS</i>	-0.1716 (0.2219)	-0.1457 (0.2892)	-0.1213 (0.3834)	-0.1000 (0.0821)	-0.4165 (0.2136)*	-0.2518 (0.2181)
<i>ENTREP</i>	-3.1042 (0.6494)***	-6.2667 (0.9300)***	-3.2844 (1.0787)***	-0.8069 (0.2427)***	-1.8054 (0.6285)***	-1.2710 (0.6339)**
<i>COMM</i>	0.0003 (0.0011)	-0.0002 (0.0007)	0.0002 (0.0015)	0.0002 (0.0002)	0.0007 (0.0005)	0.0005 (0.0006)
<i>PROV1</i>	0.9765 (0.2315)**	1.3903 (0.2794)***	1.4928 (0.3585)***	0.1722 (0.0813)**	0.4601 (0.2129)**	0.4990 (0.2271)**
<i>PROV2</i>	0.7718 (0.3630)**	1.2858 (0.4296)***	1.1754 (0.5657)**	0.2767 (0.1413)**	0.9042 (0.3897)**	0.7494 (0.3768)**
<i>PROV3</i>	1.1938 (0.4110)	1.7094 (0.4573)***	1.5731 (0.5956)***	0.2663 (0.1565)*	0.4830 (0.3286)	0.5058 (0.3468)
<i>CAPCOM</i>	0.2571 (0.2252)	0.3359 (0.1754)*	0.5373 (0.3482)	0.1034 (0.0760)***	-0.0538 (0.1028)	0.0925 (0.1443)
<i>COAST</i>	0.7423 (0.2432)***	-0.0122 (0.1156)	0.5822 (0.2615)**	0.1743 (0.0735)**	0.1721 (0.1066)*	0.2113 (0.1236)*
<i>REL</i>	0.1638 (0.0410)***	0.2910 (0.0265)***	0.2616 (0.0562)***			
<i>STU</i>				0.0092 (0.0028)***	0.0183 (0.0039)***	0.0229 (0.0063)***
Wald Test L.E.	52.46***	226.69***	53.85***	54.64***	74.602***	54.38***
Wald Test H.C.	3.34	7.37*	2.41	2.39	4.76	3.38
Wald Test Infra.	19.13***	57.96***	15.04***	19.22***	15.792***	15.85***
Wald Test Met.	3.08	20.87***	5.58	8.00	6.87	3.96
Log L	-1225.41	-1348.31	-1192.37	-608.79	-589.10	-578.15
LR Joint Test	860.34***	4400.32***	613.47***	518.03***	536.71***	269.08***
GoF Test						
LR $\alpha$ Test	527.04***			54.42***		
VuongTest		5.49***	3.56***		3.81***	3.64***

LR ZIPM-NBIM Test	311.88***	21.91***
-------------------	-----------	----------

Note: *REL* (*STU*) denotes relocations (start-ups). Marginal or partial effects of selected explanatory variables (see Panel B of Table 1) are reported. Standard errors in brackets. The symbols \*\*\*, \*\* and \* denote significance at 1%, 5% and 10% levels, respectively (906 observations). See sections 2 and 4 for definitions of the reported statistics.